Design of Takagi–Sugeno fuzzy systems by learning from examples in case a number of available data is not sufficient

Andrzej Macioł[1] , Piotr Macioł[2]

[1] AGH University of Science and Technology, Faculty of Management, Krakow
[2] AGH University of Science and Technology, Faculty of Metals Engineering and Industrial Computer Science, Krakow

Abstract: The paper describes a solution of a problem of developing of fuzzy rules compliant with the Takagi–Sugeno approach where a number of available examples (observations) is not sufficient. Modeling of fuzzy premises and generating functions describing a dependence of a result variable on antecedents are described. In our original approach a problem of identification of membership functions of variables com-posing premises and a problem of consequent parameters identification are solved. For the first one, we used a simple technique based on individual judgements of experts. The second one is solved with a linear programming method. In particular, our approach to formulate the consequent parameter identification problem allows using of an extremely effective T-S method when a data-driven approach cannot be applied. In the paper we present a description of our methods and results of simulations of accuracy of the proposed approach, based on commonly known benchmarks. The achieved accuracy of classification is sufficient for the most of decision-making systems of an expert nature.

Keywords: expert systems, fuzzy logic, knowledge acquisition, knowledge-based systems, linear programming, Takagi–Sugeno model

PROJEKTOWANIE ROZMYTYCH SYSTEMÓW TAKAGI–SUGENO METODĄ UCZENIA NA PODSTAWIE PRZYKŁADÓW, GDY LICZBA DOSTĘPNYCH DANYCH JEST NIEWYSTARCZAJĄCA

Streszczenie: W artykule opisano rozwiązanie problemu opracowania reguł rozmytych zgodnych z podejściem Takagi–Sugeno, gdy liczba dostępnych przykładów (obserwacji) jest niewystarczająca. Opisano modelowanie rozmytych przesłanek oraz generowanie funkcji opisującej zależność zmiennej wynikowej od wartości przesłanek. W naszym autorskim podejściu rozwiązany jest problem identyfikacji funkcji przynależności zmiennych definiujących przesłanki oraz problem identyfikacji parametrów funkcji opisującej konkluzje. Do rozwiązania pierwszego z nich zastosowaliśmy prostą technikę opartą na indywidualnych ocenach ekspertów. Drugi natomiast rozwiązujemy metodą programowania liniowego. W szczególności, nasze podejście do problemu identyfikacji parametrów funkcji opisującej konkluzje pozwala na zastosowanie niezwykle efektywnej metody T-S wówczas, gdy nie można zastosować podejścia opartego na danych. W pracy przedstawiamy opis naszych metod oraz wyniki symulacji dokładności proponowanego podejścia na podstawie powszechnie znanych benchmarków. Uzyskana dokładność klasyfikacji jest wystarczająca dla większości systemów decyzyjnych o charakterze eksperckim.

Słowa kluczowe: systemy ekspertowe, logika rozmyta, pozyskiwanie wiedzy, systemy oparte na wiedzy, programowanie liniowe, model Takagi–Sugeno

1. Introduction

The Fuzzy Rule-Based Systems (FRBS) is nowadays one of the most promising method of solving decision support problems, both in a form of regression tasks (Aghaeipoor and Javidi 2019) and classification tasks (i.a. Sousa et al. 2019). Although the highest expectations are currently associated with so-called black box systems, because of a transparency and an interpretability of FRBS and thanks to their ability to combine a high expressiveness with a form similar to that of experts' reasoning, they are still being developed. An important obstacle against widening of the use of FRBS is an acquisition of knowledge, especially there are no sufficient sets of training data.

In the last decades reasoning systems based on declarative knowledge acquired from observations (a data-driven approach) are opposed to those in which knowledge is formulated by experts (a knowledge-based approach) (Mutlu et al. 2017). It is generally accepted that the data-driven knowledge acquisition is much more effective (Hüllermeier 2015). This approach has been further amplified by the great interest on a knowledge discovery by machine learning – especially deep learning and data mining. Clearly, the data-driven approach requires an availability of at least a statistically significant number of patterns (observations). In some cases, a lack of the appropriate number of observations can be compensated by simulation methods. This approach can be illustrated with a reinforcement learning from games of self-play (e.g. Nakayashiki and Kaneko 2018) or metamodeling (Macioł et al. 2018).

Even if using of the data-driven approach is possible, transparency and interpretability of fuzzy systems built in this way is problematic (Chen et al. 2018). It should be also remembered that transparency and interpretability of fuzzy systems is often highlighted as one of their key advantages, especially when compared to the "black box" approximation methods, such as Artificial Neural Networks (Nasiri et al. 2013). Furthermore, an autonomous or a semi-autonomous (with a limited interaction with a human expert) defining rules from data can be also a demanding process itself (Castro et al. 2001, Ishibuchi and Yamamoto 2004).

Nevertheless, there are research problems requiring the use of an inference based on declarative knowledge, for which there is no possibility of using the data-driven knowledge acquisition or the data expansion by simulation. One can mention here, i.a.:

- designing of new technical (e.g. new materials with no sufficient physic-based numerical models), technological, organizational or business solutions etc.,
- predicting a behavior of people, communities or organizations when introducing new ways of influencing or when only fragmentary data are available,
- designing new solutions with the help of simulations, when the classical design of experiments approaches does not guarantee a representative number of examples,
- agent's behavior modeling in Multi-Agent Systems,
  and many others.

In all these cases, a discussion about ways of the data-driven knowledge acquisition is devoid of the purpose. The only available sources of knowledge are experts and/or a so-called point acquisition of knowledge method (a use of machine learning techniques for a limited number of examples/observations considered as representative by experts). Acquiring knowledge directly from experts is associated with classical "crisp" expert systems. Unfortunately, a limitation of this approach is its low expressiveness and, on the other hand, limited capabilities of experts to grasp many nuanced examples. For this approach, many learning-by-example methods have been developed (e.g. ID3), which, however, do not allow to remove the described limitations and only simplify the process of inference. Fuzzy modeling and fuzzy inference principles significantly increases the expressiveness of rule systems. One of the best-known examples is the Mamdani's method (Mamdani and Assilian 1975). The method is simple and expressive. Moreover, suitable techniques for acquiring knowledge have been developed (e.g. Wang and Mendel 1992). Results of our previous studies confirmed a significant increase in the effectiveness of this method in a relation to classical reasoning systems (Rębiasz and Macioł 2015, Macioł and Rębiasz 2018). Unfortunately, the expressiveness of the Mamdani's method and the methods of knowledge acquisition connected with it are still not sufficient to solve complex problems, especially since increasing of a number of rules may lead to high computational requirements.

The concept of Takagi and Sugeno may improve an expressiveness/computational efficiency relation both for tasks and classification tasks. Here we can also refer to the results of our research (Maciol 2017), which for a specific case showed an increase of the inference effectiveness expressed as the mean absolute percentage error by about 19% comparing to the results of the Mamdani's method (by 29% in a relation to the classical method) and by 20% comparing to the Mamdani's method (26% in a relation to the classical method) if expressed with the maximal relative deviation. Such a significant increase in the expressiveness is a result of a fact that relations between antecedents and conclusions are not described explicitly but with an arbitrary function. However, Takagi and Sugeno themselves, as well as other researchers developing their concept assumed that the formulation of such a function requires having an adequate number of training data to generate the appropriate function (by any methods and in any form).

As mentioned above, there are many problems with an insufficient number of patterns or with no patterns at all. In cases when an expert system must by developed on the ground of experts' judgments and/or "point" experiments, proper knowledge acquisition methods and techniques of reasoning, providing sufficient level of the reliability must be found.

The solution proposed in this paper addresses the situation of the insufficient number of examples (observations), while domain experts are able to provide expertise formulable as rules. The authors tried to adapt the Takagi–Sugeno (T-S) approach, initially developed for the data-driven approach, to a situation with insufficient training data.

2. Description of the problem

In the case when knowledge is defined arbitrarily by domain experts or on the basis of arbitrarily designed experiments (physical or numerical), datasets are insufficient for systematic machine learning algorithms. The choice of available alternative methods depends on chosen knowledge and reasoning models. In the case introduced above, crisp reasoning methods are the most natural choice. In this case rules formulated directly by experts or automatically extracted from given examples are in the form of Horns' clauses, which may be represented as IF... THEN rules:

IF A_i AND… AND A_n THEN B_j AND… AND B_m,

where A_i and B_j are the atoms in selected logic (propositions, predicates or other terms). In the case of "crisp" reasoning in expert systems the atoms are most often in the form of:

$$x_i \circ X \tag{1}$$

where $\circ$ is a logical connective ($=, <, >$, in, between and others) between a given value and an element of its domain (for example $x_1 = 10$, x_1 in $[2, 3, 4]$, x_1 between $<1; 10>$). The X values are selected from the domains of current input exemplary values or values ranges (called by us *support points*). For crisp reasoning, the whole process is controlled by a knowledge engineer, cooperating with a domain expert.

As it was mentioned above, the expressivity of the crisp expert systems is often insufficient. One of the possible solutions is employing fuzzy reasoning methods. The usual first-choice approach is the Mamdani model. The schema of a rule is very similar to the schema of a crisp rule and can expressed in a form

$$x_i \text{ is } X \tag{2}$$

where x_i is a crisp value of the current input and X is a linguistic term representing fuzzy set, given by a membership function $\mu_A(x)$. In this case, a knowledge acquisition process includes additionally defining of fuzzy numbers and their membership functions.

When the Takagi–Sugeno approach is employed, the form of a rule is more complex and can be expressed in a form proposed in Takagi and Sugeno (1985):

$$R : IF \left(f \left(x_1 \text{ is } A_1, \cdots, x_k \text{ is } A_k \right) \right) \text{ THEN } y = g \left(x_1, \cdots, x_k \right) \tag{3}$$

where:

y – a variable of the consequence whose value is inferred,

x_1–x_k – variables of the premises that appear also in the part of the consequence,

A_1–A_k – fuzzy sets with linear membership functions representing a fuzzy subspace in which the implication R can be applied for reasoning,

f – logical functions connecting propositions in the premise,

g – a function that implies the value of y when x_1–x_k satisfies the premise.

If the T-S is applied for a classification task, a domain of the numerical function y must be furtherly converted into a discrete set of values representing classes.

Design of a T-S model consist of four stages (Takagi and Sugeno 1985):

– identifying of variables composing premises of implications and variables representing consequences,
– formulating of rules,
– formulating of membership functions of fuzzy sets in the premises,
– establishing parameters in the consequences.

After these steps, an inference can be executed basing on the original algorithm, including all rules for each set of premises' variables.

A key problem in terms of the knowledge-based approach is solving of two problems: identifying of the membership functions of the variables composing the premises and a consequent problem of identifying of the parameters. The method proposed by us is based on a novel approach, different from those already used for the T-S methods. It differs

– in a way of formulating membership functions representing the fuzzy subspace (the linguistic variables) for the precondition's values and
– in a method of formulating a function that implies the value of the consequence.

### 2.1. The formulation of membership functions of fuzzy sets in premises in T-S approach

The first step of formulating a model of preconditions is bringing domains and hence values of all variables in premises to a common base. The essence of the T-S method is that all of input variables must be in a quantitative form. In a case where the preconditions' variables have a qualitative character, they can be converted to the ordinal quantitative one by simply projecting on a certain scale (a range variable).

In a case where we have a sufficient number of observations, the formulation of functions of fuzzy sets in premises can be implemented by various calculation methods as shown i.a. in Castro-Schez et al. (2013). In Ruspini (1970) some automatic methods of fuzzy clustering useful for this purpose are presented. Using of membership functions which are automatically derived from numerical data is presented also in in Hong and Chen (2000), Castro-Schez et al. (2004), and Fernández et al. (2015).

In Shell and Coupland (2012) a concept of a semi-automatic method for acquiring variables and their possible values is presented. Experts, who decide into how many fuzzy sets a domain of each variable will be divided, their limits and their vagueness (how much each set overlaps its neighbors) play an important role in this method. This technique is in its sense similar to our proposal. In the data-driven approach, formulating of membership functions of fuzzy sets in premises requires previous decision, how to obtain a whole Knowledge Base from a given Data Base and how to obtain a number of rules (Gacto et al. 2014).

In the absence of proper observations, the mentioned above problems become pointless. In our approach the formulation of the structure of the Knowledge Base must be implemented on the basis of subjective expert decisions. The number of rules, for example, results from the number of premises and the cardinalities of its domains. Hence, membership functions of preconditions must be determined subjectively.

In our solution the ranges of variation for the variables of the premises are pre-determined. Next, the ranges of the input variables of the premise are divided into fuzzy subspaces. The number of subspaces m_i can be different for each of variables. Each of subspaces is described by left $((c_{i,j}^-))$ and right $((c_{i,j}^+))$ *endpoints* – where i is the index of the input variable and j is the index of the subspace. The subspaces aren't contiguous (the right endpoint of the subspace i is different from the left endpoint of the subspace $i + 1$). In a complementary interval (intervals), we describe a relationship of the values of the premises to the linguistic variables as a linear function, although we do not exclude other solutions (e.g. Gaussians). In the Figure 1 we present an example of a variable which has the value between 1 and 10 projected into three linguistic values (A_1, A_2, A_3).
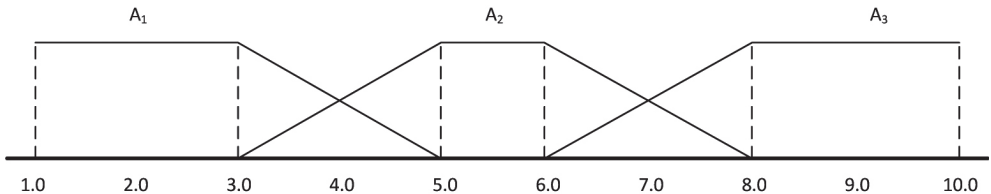
Fig. 1. The membership function of the variable projected to three linguistic values

A degree of belonging to the linguistic variables is determined by a function of any form. In the case of the knowledge structure presented above, the value of the mem-

bership function for the element x_i to the fuzzy set represented by the subset $j - \mu_j(x_i)$ is determined with an equation:

$$\mu_j(x_i) = \begin{cases} 1 & \text{if} & c_{i,j}^- \leq x_i \leq c_{i,j}^+ \\[2mm] \dfrac{c_{i,j+1}^- - x_i}{c_{i,j+1}^- - c_{i,j}^+} & \text{if } j < m \text{ and} & c_{i,j}^+ > x_i < c_{i,j+1}^- \\[4mm] \dfrac{x_i - c_{i,j-1}^+}{c_{i,j}^- - c_{i,j-1}^+} & \text{if } j > 1 \text{ and} & c_{i,j-1}^+ > x_i < c_{i,j}^- \\[4mm] 0 & & \text{elseif} \end{cases} \tag{4}$$

During the initial design of the T-S model it had been acknowledged that when ranges of conclusions of concurrently fired rules differs significantly, comparing the functions might not be straightforward. Also, a coherent method of calculation of the conclusion functions' parameters is not possible. It was decided that an internal normalization of the values of all premises is necessary. The internal normalization consists of using the highest and the lowest values of some range compared as references to transform the original scales into the range [0, 1]. In our case, we used the highest value of each premises' value as a reference. The formulation of the membership functions of the fuzzy sets in the premises had not been changed, however a subsequent normalization of both input data and boundaries of ranges describing conclusions was necessary.

2.2. The establishing of the parameters in the consequences

Due to the nature of the T-S method, the conclusion must be presented in the form of a numerical variable. Takagi and Sugeno applied their method for regression problems. In such a case, a continuous output value is a natural result. For classification tasks, numerical value of a conclusion must be transformed into a discrete range. It is straightforward in the case of control systems and the use of observations to determine the parameters in the consequences. In the case of classification tasks with the use of observations, the problem of transforming linguistic values into the numerical form and then – after reasoning – the reverse transformation appears. In the case of generating consequences basing only on experts' judgments, regardless of whether we are dealing with a problem of controlling or classification, the expert must determine a way of representation of the consequences in the numerical form. In our research, according to an approach presented i.a. in Setnes and Roubos (2000), we assumed that the value of the conclusion will be referred to as the range variable:

$$y_i \in \left\langle y_i^-, y_i^+ \right\rangle \tag{5}$$

The task of the expert is to set the limits of the ranges for each i-th linguistic value of the conclusion. Unlike the subsets of the premises, the ranges of the conclusions must be contiguous.

In the next stage, we formulate a model that allows determining a function that implies the value of the consequence according to the T-S method. In the original version of the T-S approach, one of the regression analysis methods (least squares method) is used to express the final form of the function defining the conclusion. Currently, many other methods described in the literature (Alcalá et al. 2007) allows the formulation of such a function, i.a.:

– global least squares,
– local least squares,
– product space clustering,
– evolutionary algorithms.

Since that publication, many important works in expanding and improving techniques of knowledge acquisition have been undertaken, i.a. evolutionary data base learning (Alcalá et al. 2007, Gacto et al. 2014), least squares in many modifications (Gu et al. 2017, Liu et al. 2017), convex optimization technique (Moon et al. 2019) and others. Unfortunately, all of these heuristic methods are based on an analysis of training data. To the best knowledge of the authors, there is no methods allowing formulating the function basing only on experts' predictions.

In many cases presented in the literature the problem of the lack of adequate training sets allowing to determine the parameters of the function of the conclusion is solved by using as the conclusion a constant singleton. Such a solution is presented, among others, in the case of agents modeling behavior in MAS (Francisco et al. 2019), diagnosis of diabetes (Singla 2015), industrial process control (Kaur and Kaur 2012) and many others. It is clear that this simplification significantly reduces the potential of the T-S method.

Considering the inability of using of sufficiently large training sets in our earlier studies (Maciol 2017), we used the set of equations described with the defined above *support points*. The solution of the set of equations gave coefficients in the line function sought. It turned out that for a larger number of premises, this approach does not give satisfying results. For some combinations of premises' values, the estimated value of the consequence went beyond the predicted range. To avoid this problem, an approach derived from the Linear Programming concept was used.

The core of the process of generating functions employs the support points. We assumed it would be the Cartesian product of left and right endpoints of each precondition's variables. For each combination of the values of the preconditions' variables, the consequence numerical values must be in the range set for the appropriate consequence linguistic value. For each variable describing the premise we are taking into account two extreme values and for them we are building two conditions limiting the

solution (in any case, the value of the function must not exceed the range of acceptable variability of the function describing the conclusion). In total, we obtain the number of constraints equal to double the product of all cardinalities of the domains of preconditions. Furthermore, we introduced constraints d ensuring that the extreme values of preconditions' variables will lay in some distance from boundaries of the range defined for the respective consequence (Equation (11)). We initially assumed that these limit values will be 10% of the range. During the simulation it turned out that in some cases such restriction is too sharp and do not allow finding a solution. In such cases, the restriction is iteratively weakened by one percentage point until a solution is obtained.

After solving the so-defined linear problem for each of the rules, we set the parameters of the function that implies the value of y when appropriate variables satisfy the premise:

$$y_i = p_{i,1}x_1 + p_{i,2}x_2 + \ldots + p_{i,m}x_m \tag{6}$$

During preliminary research it was decided that fitting of the conclusion functions to the extreme values is better when values of $p_{i,j}$ are small and positive. Constraining of this range might lead to the situation when the LP task is unsolvable because that arbitrarily defined constraints on the y_i variable cannot be fulfilled. The upper boundary of $p_i - u$ values is defined with an equation:

$$u = \min_i x_i \tag{7}$$

where:

$$x_i = \frac{y_i^-}{\sum\limits_{j=1}^{m} c_j^-} \tag{8}$$

That eliminates the situation when reaching of the lower boundary of the constraint y_i^- would not be possible. To eliminate the situation when obtaining the computed value lower than or equal to the upper boundary y_i^+. would not be possible, we set the values of these constraints as the minimum of the solution of the system of inequalities:

$$\sum_{j=1}^{m} uc_{i,j}^- \geq y_i^-$$
$$\sum_{j=1}^{m} uc_{i,j}^- \leq y_i^+ \tag{9}$$

and the pre-adjusted value y_i^+. If it is necessary to increase the initially assumed upper boundary (what enforces increasing of the lower boundary), it is necessary to repeat a procedure of selecting of the upper range of p_i and the upper constraint y_i^+. In a case of normalization of input variables, it is guaranteed that this problem will be solved in

a finished number of iterations. The mentioned above constraints in the LP problem for each i-th rule can be formulate as follows:

$$c_{i,1}^- p_{i,1} + c_{i,2}^- p_{i,2} + \cdots + c_{i,m-1}^- p_{i,m-1} + c_{i,m}^- p_{i,m} \leq y_i^+$$

$$c_{i,1}^- p_{i,1} + c_{i,2}^- p_{i,2} + \cdots + c_{i,m-1}^- p_{i,m-1} + c_{i,m}^- p_{i,m} \geq y_i^-$$

$$c_{i,1}^- p_{i,1} + c_{i,2}^- p_{i,2} + \cdots + c_{i,m-1}^- p_{i,m-1} + c_{i,m}^+ p_{i,m} \leq y_i^+$$

$$c_{i,1}^- p_{i,1} + c_{i,2}^- p_{i,2} + \cdots + c_{i,m-1}^- p_{i,m-1} + c_{i,m}^+ p_{i,m} \geq y_i^-$$

$$c_{i,1}^- p_{i,1} + c_{i,2}^- p_{i,2} + \cdots + c_{i,m-1}^+ p_{i,m-1} + c_{i,m}^- p_{i,m} \leq y_i^+$$

$$c_{i,1}^- p_{i,1} + c_{i,2}^- p_{i,2} + \cdots + c_{i,m-1}^+ p_{i,m-1} + c_{i,m}^- p_{i,m} \geq y_i^-$$

$$c_{i,1}^- p_{i,1} + c_{i,2}^- p_{i,2} + \cdots + c_{i,m-1}^+ p_{i,m-1} + c_{i,m}^+ p_{i,m} \leq y_i^+$$

$$c_{i,1}^- p_{i,1} + c_{i,2}^- p_{i,2} + \cdots + c_{i,m-1}^+ p_{i,m-1}$$

$$\ldots$$
(10)

$$c_{i,1}^+ p_{i,1} + c_{i,2}^+ p_{i,2} + \cdots + c_{i,m-1}^+ p_{i,m-1} + c_{i,m}^+ p_{i,m} \leq y_i^+$$

$$c_{i,1}^+ p_{i,1} + c_{i,2}^+ p_{i,2} + \cdots + c_{i,m-1}^+ p_{i,m-1} + c_{i,m}^+ p_{i,m} \geq y_i^-$$

$$\sum_{j=1}^{m} c_{i,j}^- p_{i,j} \leq \left(y_i^- + d_i \left(y_i^+ - y_i^- \right) \right)$$

$$\sum_{j=1}^{m} c_{i,j}^+ p_{i,j} \geq \left(y_i^+ - d_i \left(y_i^+ - y_i^- \right) \right)$$

$$p_{i,j} \geq 0 \ \text{ for } \ j = 1, \ldots, m$$

$$p_{i,j} \leq u \ \text{ for } \ j = 1, \ldots, m$$

where $c_{i,j}^-$ – a left endpoint of a j-th antecedent in an i-th rule and $c_{i,j}^+$ – a right endpoint of a j-th antecedent in an i-th rule.

In the proposed solution, LP is applied to find the best fit of the conclusion function to the extreme values of the premises in each rule. Hence, it is not a classical optimization problem and an objective function must be defined in a specific way. It was assumed, that a sum of distances between the conclusion functions computed for the searched parameters and the values of this function computed for the given value, constant for all premises, will be minimized. The same given value might be used for all premises thanks to normalization of premises. To eliminate a sign problem it was assumed that the constant is the value of the conclusion function computed for the

maximal values of all $p_{ij} - u$ parameters, computed as above. As a result, the objective function might be expressed in the form:

$$\sum_{k=1}^{l} \left(a_{i,k} - b_{i,k} \right) \rightarrow \min \tag{11}$$

where:

$$a_{i,1} = u(c_{i,1}^- + c_{i,2}^- + \cdots + c_{i,m-1}^- + c_{i,m}^-)$$
$$a_{i,2} = u\left(c_{i,1}^- + c_{i,2}^- + \cdots + c_{i,m-1}^- + c_{i,m}^+ \right)$$
$$\ldots$$
$$a_{i,l} = u\left(c_{i,1}^+ + c_{i,2}^+ + \cdots + c_{i,m-1}^+ + c_{i,m}^+ \right)$$

$$b_{i,1} = c_{i,1}^- p_{i,1} + c_{i,2}^- p_{i,2} + \cdots + c_{i,m-1}^- p_{i,m-1} + c_{i,m}^- p_{i,m}$$
$$b_{i,2} = c_{i,1}^- p_{i,1} + c_{i,2}^- p_{i,2} + \cdots + c_{i,m-1}^- p_{i,m-1} + c_{i,m}^+ p_{i,m}$$
$$\ldots$$
$$b_{i,l} = c_{i,1}^+ p_{i,1} + c_{i,2}^+ p_{i,2} + \cdots + c_{i,m-1}^+ p_{i,m-1} + c_{i,m}^+ p_{i,m}$$
$$l = 2^m$$

$$\tag{12}$$

A reasoning in our approach may be conducted identically to the classical T-S approach. In our research, we accepted the operator PROD as an aggregation function for rules. Tests using other operators had not bring visible changes in simulation results. Hence, a result is a continuous numerical value. If the consequences function is defined only basing on experts' judgements and the reasoning is used for classification, the expert himself must define how the numerical value of the consequence is interpreted linguistically. We assumed that the value of the conclusion is within the particular range and after reasoning is transformed to the linguistic form. The value of conclusions is calculated in following form:

$$result = \frac{\sum_{k}^{l} w_k y_i}{\sum_{k}^{l} w_k} \tag{13}$$

where:

$$w_k = \prod_{j=1}^{m} \frac{x_j - c_{i,j}^-}{c_{i,j}^+ - c_{i,j}},$$

x_j – a current value of the j-th precondition,
k – an index of the fired rule,
l – a number of the fired rule.

The rule is fired when all values of the analyzed case's preconditions are within the range set for this rule or are close to its boundaries (they are in the ranges of "neighboring" rules).

In earliest publications authors used a simple mechanism, dividing the range to equal subspaces (Setnes and Roubos 2000). Each subspace corresponds to the particular linguistic value. This approach is referred below as "classical".

In our opinion this solution can be used for a limited cardinality of domains of conclusions and with a monotonous dependence of the value of the conclusions on the value of premises. We have examined a possibility of using other methods of transforming numerical result of reasoning into linguistic values. We proposed that a "similarity index" rather than an average value should be decisive in classifying the conclusions value into the certain class. Such an approach assigns the numerical result of reasoning of y_i to the particular class (defined with the interval constant) with a function, defining a degree of membership into the particular range with the Manhattan distance to the boundary of the range. We studied two variants of classification. In the first one, regardless of the conclusion originally assigned to the given rule, the range in which the inference result was located was sought (variant I). We also assumed that for the extreme values of the ranges of the conclusions, the distance will be measured from the lower or upper boundary of the range, respectively. The "similarity index" s_i replacing a directly computed value y_i is set according to the equation:

$$
s_i = \begin{cases}
1 - \dfrac{y_i}{c_1^+ - c_1^-} & \text{if } y_i \leq c_1^+ \\[3mm]
1 - \dfrac{ABS\left(c_i^- + \dfrac{1}{2}\left(c_i^+ - c_i^-\right) - y_i\right)}{\dfrac{1}{2}\left(c_i^+ - c_i^-\right)} & \text{if } c_i^- \leq y_i \leq c_i^+ \text{ and } 1 < i < n \\[3mm]
\dfrac{y_i - c_{cn}^i}{c_{cn}^+ - c_{cn}^-} & \text{if } y_i \geq c_{cn}^-
\end{cases}
\tag{14}
$$

where cn – a number of the conclusions values. The ultimate classification proceeds the same way as described above.

In the second option, only the range originally assigned to the conclusion of the rule was taken into account (variant II). For each conclusion of the fired rule y_i,

the distances from the lower (*dl*) and the upper (*du*) boundaries of the conclusion's range for this rule is computed. These values are computed with the equation:

$$dl = \frac{y_i - c_i^-}{c_i^+ - c_i^-} \text{ if } c_i^- \le y_i \le c_i^+$$

$$dl = 1 \text{ if } y_i > c_i^+$$

$$dl = 0 \text{ if } y_i < c_i^-$$

$$du = \frac{c_i^+ - y_i}{c_i^+ - c_i^-} \text{ if } c_i^- \le y_i \le c_i^+$$

$$du = 1 \text{ if } y_i < c_i^-$$

$$du = 0 \text{ if } y_i > c_i^+$$

(15)

These distances are weighted (as in the classical variant) and summed if more than one rule points to the same conclusion. The classification proceeds according to the following algorithm:

```
Flag = True
MaxResult = 0
For i = 1 to cn
    If du(i) <> 0 And Flag Then
        Flag = False
        MaxResult = du(i)
        Result = i
    Else
        If dl(i) > MaxResult Then
            MaxResult = dl(i)
            Result = i
        End If
    End If
Next
```

3. Experimental results

The assumption that knowledge is formulated by experts basing on their subjective judgments and is not based on a set of examples (observations) precludes in principle a possibility of verifying a correctness of so-formulated rules. However, we have made an attempt to evaluate the accuracy of the proposed inference method with the commonly used benchmarks.

Our goal was verifying the hypothesis, that the T-S method might be successfully used for solving classification tasks with knowledge acquired from human experts, not from data. In our work we also have examined results obtained with the classical "crisp" reasoning methods and the Mamdani's method, with the same set of rules as in the proposed T-S approach modification. In these approaches knowledge acquisition is simpler, however the authors' earlier research suggests worse effectiveness of classification.

3.1. Example 1

Firstly, we used the iris data commonly used in classification and pattern recognition studies (Fisher 1936). It contains 150 measurements of four features from three species: *Iris setosa*, *Iris versicolor* and *Iris virginica*. The features are: sepal length – x_1, sepal width – x_2, petal length – x_3, and petal width – x_4. Our experiment consisted of formulating the rules based on data presented in the iris data set, but without using any data analysis methods. It was assumed that all of features will be described with the scale [small, large].

The initial division of the premises into two categories resulted with 16 rules corresponding to the Cartesian product of the four features (each of two values). For each of the generated sets of premises we set the expected value of the answer (the species) using of course the examples but not using any analytical methods. The rules defined in this way are presented in Table 1. Basing on the analysis of the examples, assuming in principle a regular distribution, we have established the ranges of the linguistic values for each feature (Tab. 2).

Table 1
The rules for iris data

Rule No.	Premises				Consequence
	x_1	x_2	x_3	x_4	
1	small	small	small	small	*Iris setosa*
2	small	small	small	large	*Iris versicolor*
3	small	small	large	small	*Iris versicolor*
4	small	small	large	large	*Iris virginica*
5	small	large	small	small	*Iris setosa*
6	small	large	small	large	*Iris versicolor*
7	small	large	large	small	*Iris setosa*
8	small	large	large	large	*Iris versicolor*
9	large	small	small	small	*Iris versicolor*
10	large	small	small	large	*Iris virginica*

Table 1 cont.

11	large	small	large	small	*Iris versicolor*
12	large	small	large	large	*Iris virginica*
13	large	large	small	small	*Iris versicolor*
14	large	large	small	large	*Iris versicolor*
15	large	large	large	small	*Iris versicolor*
16	large	large	large	large	*Iris virginica*

Table 2
Boundaries limits for features

		Feature								
		x_1		x_2		x_3		x_4		
		source	normalized	source	normalized	source	normalized	source	normalized
Range:		4.30–7.90	0.54–1.00	2.00–4.40	0.45–1.00	1.00–6.90	0.14–1.00	0.10–2.50	0.04–1.00
small	–	4.30	0.54	2.00	0.45	1.00	0.14	0.10	0.04
	+	5.74	0.73	2.96	0.67	3.36	0.38	1.06	0.40
large	–	6.46	0.82	3.44	0.78	4.54	0.66	1.54	0.68
	+	7.90	1.00	4.40	1.00	6.90	1.00	2.50	1.00

To perform the classification, the output of the T-S model was used with the following classification rule:

$$cl = \begin{cases} 1, & \text{if } 0.0 \le y_k < 1.5 \\ 2, & \text{if } 1.0 \le y_k < 3.0 \\ 3, & \text{if } 2.0 \le y_k < 4.5 \end{cases} \tag{16}$$

where: cl – the classifier, 1 – *Iris setosa*, 2 – *Iris versicolor*, 3 – *Iris virginica*.

Let consider the example of the first rule:

```
IF sepal length is small
AND sepal width is small
AND petal length is small
AND petal width is small
THEN species is Iris-setosa
```

The adopted rules allow to formulate an LP task:

– The constraints arising from the required inclusion of the conclusion's value within the given boundaries:

$$0.54 \cdot p_{1,1} + 0.4 \cdot p_{1,2} + 0.14 \cdot p_{1,3} + 0.04 \cdot p_{1,4} \leq 1.5$$

$$0.54 \cdot p_{1,1} + 0.45 \cdot p_{1,2} + 0.14 \cdot p_{1,3} + 0.04 \cdot p_{1,4} \geq 0$$

$$0.54 \cdot p_{1,1} + 0.45 \cdot p_{1,2} + 0.14 \cdot p_{1,3} + 0.40 \cdot p_{1,4} \leq 1.5$$

$$0.54 \cdot p_{1,1} + 0.45 \cdot p_{1,2} + 0.14 \cdot p_{1,3} + 0.40 \cdot p_{1,4} \geq 0$$

$$0.54 \cdot p_{1,1} + 0.45 \cdot p_{1,2} + 0.38 \cdot p_{1,3} + 0.04 \cdot p_{1,4} \leq 1.5$$

$$0.54 \cdot p_{1,1} + 0.45 \cdot p_{1,2} + 0.38 \cdot p_{1,3} + 0.04 \cdot p_{1,4} \geq 0 \qquad (17)$$

$$0.54 \cdot p_{1,1} + 0.45 \cdot p_{1,2} + 0.38 \cdot p_{1,3} + 0.40 \cdot p_{1,4} \leq 1.5$$

$$0.54 \cdot p_{1,1} + 0.45 \cdot p_{1,2} + 0.38 \cdot p_{1,3} + 0.40 \cdot p_{1,4} \geq 0$$

…

$$0.73 \cdot p_{1,1} + 0.67 \cdot p_{1,2} + 0.38 \cdot p_{1,3} + 0.40 \cdot p_{1,4} \leq 1.5$$

$$0.73 \cdot p_{1,1} + 0.67 \cdot p_{1,2} + 0.38 \cdot p_{1,3} + 0.40 \cdot p_{1,4} \geq 0$$

– The constraints enforcing to the extreme values of preconditions' variables the total maximum distance of the consequence value from its boundaries to be greater than a given value (10%):

$$0.54 \cdot p_{1,1} + 0.45 \cdot p_{1,2} + 0.14 \cdot p_{1,3} + 0.04 \cdot p_{1,4} \leq 0.15$$
$$0.73 \cdot p_{1,1} + 0.67 \cdot p_{1,2} + 0.38 \cdot p_{1,3} + 0.40 \cdot p_{1,4} \geq 1.35 \qquad (18)$$

– The additional constraints:

$$p_{1,j} \geq 0 \ \text{ for } j = 1, 2, 3, 4$$
$$p_{1,j} \leq 1.6 \ \text{ for } j = 1, 2, 3, 4 \qquad (19)$$

– The objective function:

$$\left(1.5 \cdot (0.54 + 0.45 + 0.14 + 0.04) - \left(0.54 \cdot p_{1,1} + 0.45 \cdot p_{1,2} + 0.14 \cdot p_{1,3} + 0.04 \cdot p_{1,4}\right)\right) +$$

$$+ \left(1.5 \cdot (0.54 + 0.45 + 0.14 + 0.40) - \left(0.54 \cdot p_{1,1} + 0.45 \cdot p_{1,2} + 0.14 \cdot p_{1,3} + 0.40 \cdot p_{1,4}\right)\right) +$$

$$+ \cdots + \qquad (20)$$

$$+ \left(1.5 \cdot (0.73 + 0.67 + 0.38 + 0.40) - \left(0.73 \cdot p_{1,1} + 0.67 \cdot p_{1,2} + 0.38 \cdot p_{1,3} + 0.40 \cdot p_{1,4}\right)\right) \rightarrow$$

$$\rightarrow \min$$

After solving the LP problem, we get the following rule:

```
IF sepal length is small
AND sepal width is small
AND petal length is small
AND petal width is small
THEN y₁ = 0.0000 · x₁ + 0.3797 · x₂ + 1.5000 · x₃ + 1.5000 · x₄
```

The next 15 rules were generated in the same way.

The accuracy of the knowledge acquisition approach was tested by comparing the classification made with the T-S algorithm with the real values. The simulations were carried out for three mentioned above methods of transforming of the numerical values into the linguistic values (classical, variant I and variant II). Differences between the mechanisms of transformation can be shown with the example. If $x_1 = 6.8$, $x_2 = 2.8, x_3 = 4.8$ and $x_4 = 1.4$ (when normalized respectively 0.86, 0.64, 0.70 and 0.56) then the valid classification is iris versicolor. For this case the rules presented in the Table 3 are used.

Table 3
The rules used in an example of reasoning

Rule No.	Premises				Consequence	Reasoning result	Weight	Weighted result
11	large	small	large	small	*Iris versicolor*	2.7277	0.4286	1.1691
	0.0000	1.3266	1.5000	1.5000				
12	small	large	large	large	*Iris virginica*	3.3224	0.5714	1.8984
	0.8652	1.5000	1.1257	1.5000				
Total								3.0675

According to the classical method used in the T-S classifiers, the linguistic conclusion is *Iris virginica* because the numerical result of reasoning (3.0675) is within the range <3.0, 4.5>. For the variant II, the same result is obtained. The relative distance of the numerical result of the rule 11 from the boundary between ranges indicating for iris-versicolor and iris-virginica (3.0) is smaller than the (relative) distance between the numerical result of the rule 12 and this boundary. A different transformation result may be obtained with the variant I because the numerical result of reasoning with the rule 11 closer to the middle of the range <1.5, 3.0> then the result of rule 12 to upper bound of range <3.0, 4.5>. The rule 11 indicates *Iris versicolor*. All three methods of classification are shown in Figure 2.
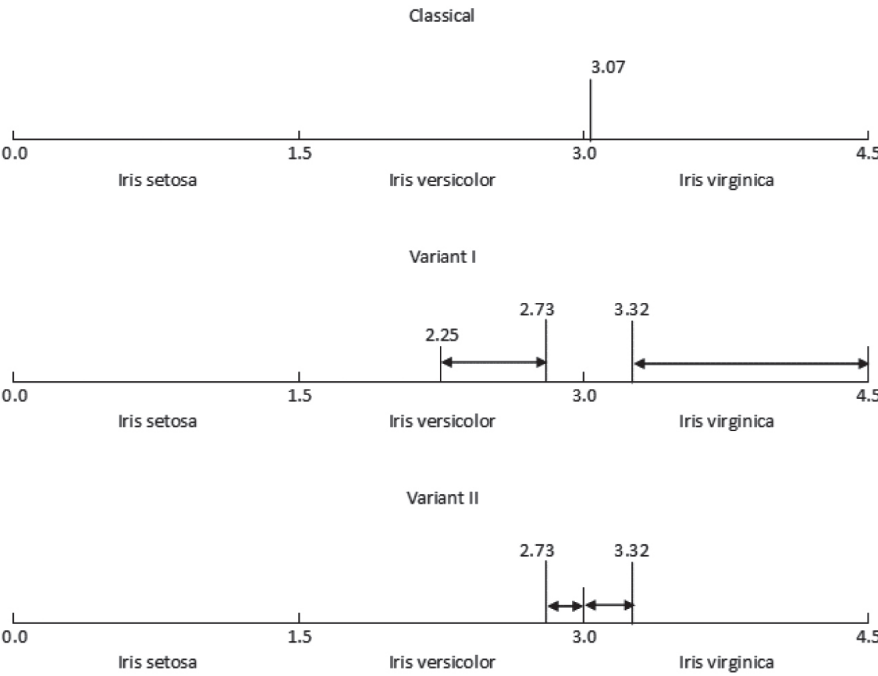
Fig. 2. Comparison of results with three different approaches to classification

The simulation examined all three methods. In each case, the accuracy of the classification was measured as the ratio of the number of correctly classified objects of the given set to the number of all objects of this set. The results of the simulation are presented in Table 4.

Table 4
The results of the simulation for iris data

Method	Effectiveness of classification [%]
Classical	91.33
Variant I	92.67
Variant II	92.67

The presented results are comparable with those obtained for the same data set using different methods based on large sets of learning examples. Riza et al. (2015) presented results of classification with several fuzzy rule-based classification methods. The effectiveness was between 94.00 and 97.34%.

Classification with the same rules, but with the crisp reasoning and the Mamdani's approach resulted with significantly worse effectiveness, respectively 82.00 and 87.33%.

3.2. Example 2

The simulation of the commonly used cancer database obtained from the University of Wisconsin Hospitals, Madison by Wolberg et al. (1995) was also provided. In the original version, the database includes 699 samples with nine attributes which were primarily used to classify breast cancer as either benign or malignant basing on a cell description gathered by a microscopic examination. Each attribute is described on the scale of 1 to 10. Such a large number of attributes excludes the possibility of using an expert judgment even after transforming a numerical scale into a qualitative scale (in the case of two values of the linguistic variables, we have $2^9 = 512$ examples). However, studies based on the data-driven approach (Garg et al. 2009) shows that only three attributes: uniformity of cell size (x_1), single epithelial cell size (x_2) and Normal Nucleoli (x_3) attributes are significantly important to decide the tumor type. Limiting the number of attributes allowed us to use the mentioned data set to verify our concept. We transformed data in the numerical scale of 1 to 10 to the qualitative values [small, large]. Basing on sample data but like in the Example 1 without using any data analysis methods, we have formulated eight rules. Three simulations analogous to those carried out in Example 1 were conducted. The results are presented in Table 5.

Table 5
The results of the simulation for cancer data

Method	Effectiveness of classification [%]
Classical	89.41
Variant I	89.41
Variant II	94.85

In this case the accuracy of classification is slightly lower than results presented in other studies. For example in Bekaddour (2012), the results of classification with the neuro-fuzzy inference technique ANFIS (Adaptative Neuro-Fuzzy Inference System) are presented. The results show that the best performances are obtained by this model compared to others cited in literature (98.25%). The authors believe that a superiority of the data-based approach arises from much higher number of samples than in Example 1.

As expected, the "crisp" reasoning resulted with the lower accuracy, equal to 90.56%. A surprisingly good result was obtained with the classical Mamdani method (95.14%). This result is better the one obtained with the T-S method. In the authors opinion, that might be caused by the fact that both input data and classifier have a dichotomous form. It significantly reduces advantages of T-S method. However, proving this hypothesis requires further works.

3.3. Example 3

We have attempted to verify the results of our method in comparison to data-driven methods in the case when the number of antecedents in reasoning rules is too large for them to be covered by experts, but there is a possibility of "structuring" of the knowledge base (Mutlu et al. 2017). As a test case we used the so-called "Adults" data set (Kohavi and Becker 1996). The US Adult Census dataset is a repository of 48 842 entries extracted from the 1994 US Census database. The database is used to form models to predict whether an individual made more or less than $50 000 in 1994. Each entry contains 15 variables referring to an individual. These attributes have qualitative or quantitative character. Such a large number of attributes and cardinality of its domains excludes the possibility of using a knowledge-based approach to describe the problem even after the transformation of continuous values into linguistic variables.

Due to the purpose of the research (verification of the effectiveness of the proposed method), the Adults problem has been simplified. The number of examples was limited to the group "Male" and "White" and those for whom there were no cases of capital-gain and capital-loss. The corresponding attributes have been eliminated. We removed the continuous attribute *fnlwgt* (final weight). We also eliminated attribute *education* because it is represented by the *education-num*. We have limited the domain of the *marital-status* attribute to values: Married-civ-spouse, Divorced, Never-married, Separated, eliminating Married-AF-spouse, Married-spouse-absent and Widowed due to the small number of suitable examples. We have reduced the domain of the *occupation* attribute to values Tech-support, Craft-repair, Other-service, Sales, Exec-managerial, Prof-specialty, Handlers-cleaners, Machine-op-inspct, Adm-clerical, Farming-fishing, Transport-moving, Protective-serv, eliminating Armed-Forces and Priv-house-serv for the same reasons.

As a result of these activities, we obtained 13 988 entries containing six attributes. The number of examples to consider due to the multiplicity of the Cartesian product of attribute domains was still too high (216).

At the same time, it was found that the attributes can be rationally divided into two clearly separated groups.

Personal data:
– Age cut into levels Young (0–30), Middle-aged (40–55) and Senior (65+).
– Education-num: the highest level of education achieved in numerical form transformed into 3 linguistic values: Low (1–3.5), Middle (7.5–10), High (14–16).
– Marital-status: marital status of an individual transformed into 3 numerical classes 1 (Never-married), 2 (Divorced and Separated), 3 (Married-civ-spouse) and then two linguistic variables Low (1) and High (3) – class 2 is the intermediate value between Low and High.

Professional data:

– Work-class: a general term to represent the employment status of an individual transformed into four numerical classes: 1 (Private and Self-emp-not-inc), 2 (Local-gov and State-gov), 3 (Federal-gov), 4 (Self-emp-inc) and then two linguistic variables Low (1) and High (4) – classes 2 and 3 are the intermediate value between Low and High.

– Occupation: the general type of occupation of an individual transformed into four numerical classes: 1 (Farming-fishing, Handlers-cleaners, Machine-op--inspct and Other-service), 2 (Adm-clerical, Craft-repair and Transport-moving), 3 (Protective-serv, Sales and Tech-support), 4 (Exec-managerial and Prof-specialty) and then two linguistic variables Low (1) and High (4) – classes 2 and 3 are the intermediate value between Low and High.

– Hours-per-week: the hours an individual has reported two work per week cut into levels Small (0–38), Middle (42–48) and Large (52+).

In the rules describing personal data in the numerical form, we used as consequences a scale of six steps (1st is the lowest chance of earning more than 50K and 6th largest). In case of the rules describing professional data, we used a scale of four steps (1st is the lowest chance of earning more than 50K and 4th largest).

In the case of "structuring" the knowledge base, reasoning is carried out in the subsequent stages allowing for determining further intermediate values. In the presented case, in the first stage, the values of linguistic variables defined as personal and professional and then the value of the result variable (below 50K, above 50K) are determined. Hence, we have three decision problems for which rules of inference have been established on principles similar to those used in previous examples.

We perform three simulations analogously to those carried out in examples 1 and 2. The results are presented in Table 6.

Table 6
The results of the simulation for "Adults" data

Method	Effectiveness of classification [%]
Classical	76.89
Variant I	76.29
Variant II	76.16

The obtained results differ significantly from the newest results obtained by data-mining methods. For example Chakrabarty and Biswas (2018) presented results of classification with the Gradient Boosting Classifier model and the accuracy of 88.16%.

That result is significantly better than previous works. So big difference between the data-driven methods and results obtained with experts' knowledge arises, just like in the Example 2, from availability of a large number of training samples.

Again, results were compared with the ones obtained with "crisp" reasoning and Mamdani's method. In both cases, results were worse, respectively 75.16 and 75.26%.

3.4. Example 4

We have also attempted to verify the results of the proposed method with an example of Car Evaluation Database. This sample is a collection of the records (1728) on specific attributes on cars donated by Bohanec in 1997 and was obtained from the UCI dataset repository. This evaluates cars according to the following concept structure:

– CAR – car acceptability,
– PRICE – overall price,
– BUYING – buying price,
– MAINT – price of the maintenance,
– TECH – technical characteristics,
– COMFORT – comfort,
– DOORS – number of doors,
– PERSONS – capacity in terms of persons to carry,
– LUG_BOOT – the size of luggage boot,
– SAFETY – estimated safety of the car.

In the earlier researches that problem was treated as a classical structural task. Below, other approach is presented. The presented method was used with 64 examples (rules). For six previously mentioned variables (BUYING, MAINT, DOORS, PERSONS, LUG_BOOT, SAFETY – each one having values small or large).

Table 7
The results of the simulation for "cars" data

Method	Effectiveness of classification [%]
Classical	83.68
Variant I	82.70
Variant II	82.41

As previously, the results are significantly worse than the ones obtained with the data-driven learning. Effectivity presented in Awwalu et al. (2014) varies between 90,81 and 93,51%, depending on validation method. To improve the accuracy, the

authors had tried to introduce a hierarchical model into presented approach. Unfortunately, results were not satisfying. As previously, the "crisp" reasoning and the Mamdani's method were also applied. In both cases, obtained accuracy was equal to 74.65%, what is noticeably worse than when the same rules were used for reasoning with the T-S method.

3.5. Summary of the results

Results of four examples and three methods of aggregating simulation are compared with the results obtained with the data-driven, the "crisp" reasoning and the Mamdani's methods in Figure 3 and Table 8.
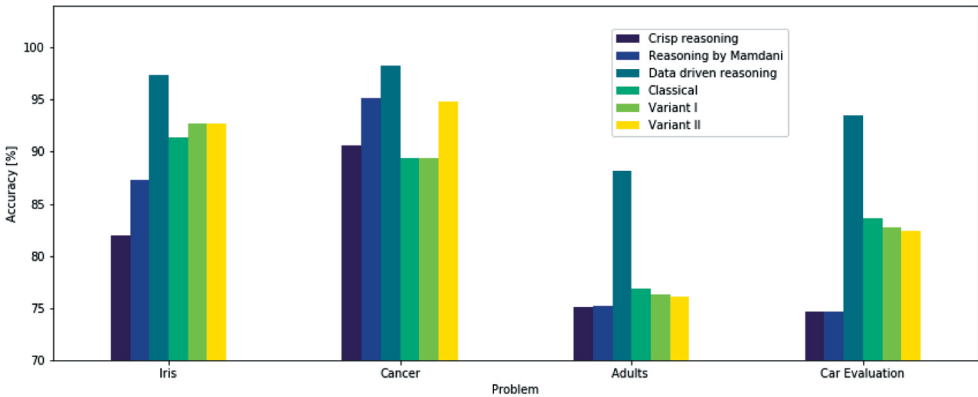

Fig. 3. Efficiency of classification for Iris, Cancer, Adults and Cars datasets

Table 8
The results of the simulation – summary

Problem	Method			
	Crisp reasoning [%]	Reasoning by Mamdani [%]	Data-driven reasoning [%]	The best result of our attempt [%]
Iris	82.00	87.33	97.34	92.67
Cancer	90.56	95.14	98.25	94.85
Adults	75.16	75.26	88.16	76.89
Car evaluation	74.65	74.65	93.51 (by 10-folds cross validation)	83.68

In each of the presented examples, the classification efficiency is relatively high and comparable to the results obtained using data-driven methods. The results of

simulations for different methods of transforming the numerical results of reasoning into classifiers indicate relatively small differences when using different techniques. In three cases out of four, the classification results obtained with our approach were better than the ones obtained with the Mamdani's method and in all cases were better than in the "crisp" reasoning.

4. Conclusions

The presented research was aimed at verifying of the hypothesis, is employing the T-S approach for reasoning possible when the knowledge base is built with expert's judgments and/or "local" experiments and there is no possibility to use large training sets. The results presented above, in the authors opinion, proves that the hypothesis is true

In our opinion, the achieved accuracy of classification is sufficient due to the needs of decision-making systems of an expert nature. The T-S method together with the presented approach of acquiring knowledge creates completely new possibilities of using the idea of artificial intelligence in those areas in which, for obvious reasons, it is impossible to use adequately numerous sets of examples or observations. It is surprising that in one case (iris example) the simulation's results obtained by us outweigh the efficiency of classification obtained when using commonly used data-driven methods. It may indicate that in some cases, even when data is available, but their preparation and processing is "expensive", it may be justified to use the proposed method. Furthermore, the obtained results show that in some situations equally good effectiveness might be achieved with Mamdani's reasoning method.

Confirmation of the correctness of our findings requires further research. The easiest way is to simulate subsequent known benchmarks. The results of these simulations will also help to solve the problem of selecting the best method of transforming the numerical results of inference into classification results.

We intend to focus our research on solving technological problems and in particular on planning laboratory and/or numerical experiments in the case when traditional methods fail. The aim will be to answer the question whether using certain qualitative model of technological problems we can effectively determine the previously mentioned support points. We expect that it will be possible to describe the variability of parameters describing certain phenomena and then applying our method to obtain a sufficiently reliable model of the studied phenomenon. Our predictions have already been tentatively confirmed in studies on prediction of forging dies wear (Macioł et al. 2020) and microstructure evolution prediction of the metallic products.

References

Aghaeipoor F., Javidi M.M., 2019, *MOKBL+MOMs: An interpretable multi-objective evolutionary fuzzy system for learning high-dimensional regression data*, Information Sciences, vol. 496, pp. 1–24. https://doi.org/10.1016/j.ins.2019.04.035.

Alcalá R., Alcalá-Fdez J., Casillas J., Cordón O., Herrera F., 2007, *Local identification of prototypes for genetic learning of accurate TSK fuzzy rule-based systems*, International Journal of Intelligent Systems, vol. 22, pp. 909–941. https://doi.org/10.1002/int.20232.

Awwalu J., Ghazvini A., Abu Bakar A., 2014, *Performance Comparison of Data Mining Algorithms: A Case Study on Car Evaluation Dataset*, International Journal of Computer Trends and Technology, vol., pp. 78–82. https://doi.org/10.14445/22312803/IJCTT-V13P117.

Bekaddour F., 2012, *A Neuro-Fuzzy Inference Model for Breast Cancer Recognitio*, International Journal of Computer Science and Information Technology, vol. 4, pp. 163–173. https://doi.org/10.5121/ijcsit.2012.4513.

Castro-Schez J.J., Castro J.L., Zurita J.M., 2004, *Fuzzy repertory table: A method for acquiring knowledge about input variables to machine learning algorithm*, IEEE Transactions on Fuzzy Systems, vol. 12, pp. 123–139. https://doi.org/10.1109/TFUZZ.2003.822684.

Castro-Sche, J.J., Murill, J.M., Migue, R., Lu, X., 2013, *Knowledge acquisition based on learning of maximal structure fuzzy rules*, Knowledge-Based Systems, vol. 44, pp. 112–120. https://doi.org/10.1016/j.knosys.2013.01.033.

Castro J.L., Castro-Sanchez J.J., Zurita J.M., 2001, *Use of a fuzzy machine learning technique in the knowledge acquisition process*, Fuzzy Sets and Systems, vol. 123, pp. 307–320. https://doi.org/10.1016/S0165-0114(01)00008-2.

Chakrabarty N., Biswas S., 2018, *A Statistical Approach to Adult Census Income Level Prediction*, [in:] *2018 International Conference on Advances in Computing, Communication Control and Networking (ICACCCN)*, pp. 207–212. https://doi.org/10.1109/ICACCCN.2018.8748528.

Chen T., Shang C., Su P., Shen Q., 2018, *Induction of accurate and interpretable fuzzy rules from preliminary crisp representation*, Knowledge-Based Systems, vol. 146, pp. 152–166.

Fernández A., López V., del Jesus M.J., Herrera F., 2015, *Revisiting Evolutionary Fuzzy Systems: Taxonomy, applications, new trends and challenges*, Knowledge-Based Systems, vol. 80, pp. 109–121. https://doi.org/10.1016/j.knosys.2015.01.013.

Fisher R.A., 1936, *The use of multiple measurements in taxonomic problems*, Annals of Eugenics, vol. *7*(2), pp. 179–188. https://doi.org/10.1111/j.1469-1809.1936.tb02137.x.

Francisco M., Mezquita Y., Revollar S., Vega P., De Paz J.F., 2019, *Multi-agent distributed model predictive control with fuzzy negotiation*, Expert Systems with Applications, vol. 129, pp. 68–83. https://doi.org/10.1016/j.eswa.2019.03.056.

Gacto M.J., Galende M., Alcalá R., Herrera F., 2014, *METSK-HDe: A multiobjective evolutionary algorithm to learn accurate TSK-fuzzy systems in high-dimensional and large-scale regression problems*, Information Sciences, vol. 276, pp. 63–79. https://doi.org/10.1016/j.ins.2014.02.047.

Garg B., Sufian Beg M.M., Ansari A.Q., 2009, *Optimizing Number of Inputs to Classify Breast Cancer Using Artificial Neural Network*, Journal of Computer Science & Systems Biology, vol. 2(4), pp. 247–254. https://doi.org/10.4172/jcsb.1000037.

Gu X., Chung F., Ishibuchi H., Wang S., 2017, *Imbalanced TSK Fuzzy Classifier by Cross-Class Bayesian Fuzzy Clustering and Imbalance Learning*, EEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 47, no. 8, pp. 2005–2020. https://doi.org/10.1109/TSMC.2016.2598270.

Hong T.-P., Chen J.-B., 2000, *Processing individual fuzzy attributes for fuzzy rule induction*, Fuzzy Sets and Systems, vol. 112(1), pp. 127–140. https://doi.org/10.1016/S0165-0114(98)00179-1.

Hüllermeier E., 2015, *From knowledge-based to data-driven fuzzy modeling*, Informatik-Spektrum, vol. 38, pp. 500–509. https://doi.org/10.1007/s00287-015-0931-8.

Ishibuchi H., Yamamoto T., 2004, *Fuzzy rule selection by multi-objective genetic local search algorithms and rule evaluation measures in data mining*, Fuzzy Sets and Systems, vol. 141, pp. 59–88. https://doi.org/10.1016/S0165-0114(03)00114-3.

Kaur A., Kaur A., 2012, *Comparison of Mamdani-Type and Sugeno-Type Fuzzy Inference Systems for Air Conditioning System*, International Journal of Soft Computing and Engineering, vol. 2, pp. 323–325.

Kohavi R., Becker B., 1996, *Adult Data Set*, https://archive.ics.uci.edu/ml/datasets/adult [access: 5.11.2018].

Liu J., Chung F., Wang S., 2017, *Bayesian zero-order TSK fuzzy system modeling*, Applied Soft Computing, vol. 55, pp. 253–264. https://doi.org/10.1016/j.asoc.2017.01.040.

Maciol A., 2017, *Knowledge-based methods for cost estimation of metal casts*, The International Journal of Advanced Manufacturing Technology, vol. 91, pp. 641–656. https://doi.org/10.1007/s00170-016-9704-z.

Macioł A., Rębiasz B., 2018, *Multicriteria Decision Analysis (MCDA) Methods in Life Cycle Assessment (LCA): A Comparison of Private Passenger Vehicles*, Operations Research and Decisions, vol. 28, pp. 5–26. https://doi.org/10.5277/ord180101.

Macioł P., Szeliga D., Sztangret Ł., 2018, *Methodology for metamodelling of microstructure evolution: precipitation kinetic case study*, International Journal of Material Forming, vol. 11, pp. 867–878. https://doi.org/10.1007/s12289-017-1396-x.

Macioł A., Macioł P., Mrzygłód B., 2020, *Prediction of forging dies wear with the modified Takagi-Sugeno fuzzy identification method*, Materials and Manufacturing Processes, vol. 35(6), pp. 700–713. https://doi.org/10.1080/10426914.2020.1747627.

Mamdani E.H., Assilian S., 1975, *An experiment in linguistic synthesis with a fuzzy logic controller*, International Journal of Man-Machine Studies, vol. 7, pp. 1–13. https://doi.org/10.1016/S0020-7373(75)80002-2.

Moon H., Kim A.-H., Joo Y.H., Bae J., 2019, *A Novel Approach to Design of Takagi–Sugeno Fuzzy Classifier*, https://www.researchgate.net/publication/228405644_A_Novel_Approach_to_Design_of_Takagi-Sugeno_Fuzzy_Classifier [access: 12.12.2018].

Mutlu B., Sezer E.A., Nefeslioglu H.A., 2017, *A defuzzification-free hierarchical fuzzy system (DF-HFS): Rock mass rating prediction*, Fuzzy Sets and Systems, vol. 307, pp. 50–66. https://doi.org/10.1016/j.fss.2016.01.001.

Nakayashiki T., Kaneko T., 2018, *Learning of Evaluation Functions via Self-Play Enhanced by Checkmate Search*, [in:] *2018 Conference on Technologies and Applications of Artificial Intelligence (TAAI)*, pp. 126–131. https://doi.org/10.1109/TAAI.2018.00036.

Nasiri M., Fober T., Senge R., Hullermeier E., 2013, *Fuzzy pattern trees as an alternative to rule-based fuzzy systems: Knowledge-driven, data-driven and hybrid modeling of color yield in polyester dyeing*, [in:] *2013 Joint IFSA World Congress and NAFIPS Annual Meeting (IFSA/NAFIPS)*, IFSA/NAFIPS, pp. 715–721. https://doi.org/10.1109/IFSA-NAFIPS.2013.6608488.

Rębiasz B., Macioł A., 2015, *Comparison of Classical Multi-Criteria Decision Making Methods with Fuzzy Rule-Based Methods on the Example of Investment Projects Evaluation*, [in:] Neves-Silva R., Jain L.C., Howlett R.J. (eds.), *Intelligent Decision Technologies: proceedings of the 7th KES international conference on Intelligent Decision Technologies (KES-IDT 2015)*, Springer International Publishing, Cham, pp. 549–561.

Riza L.S., Bergmeir C., Herrera F., Benitez J.M., 2015, *FRBS: Fuzzy rule-based systems for classification and regression in R*, Journal of Statistical Software, vol. 65, pp. 1–30. https://doi.org/10.18637/jss.v065.i06.

Ruspini E.H., 1970, *Numerical Methods for Fuzzy Clustering*, Information Sciences, vol. 2, pp. 319–350. https://doi.org/10.1016/S0020-0255(70)80056-1.

Setnes M., Roubos H., 2000, GA-fuzzy modeling and classification: complexity and performance, IEEE Transactions on Fuzzy Systems, vol. 8, no. 5, pp. 509–522. https://doi.org/10.1109/91.873575.

Shell J., Coupland S., 2012, *Towards Fuzzy Transfer Learning for Intelligent Environments*, [in:] Paternò F., de Ruyter B., Markopoulos P., Santoro C., van Loenen E., Luyten K. (ed.), *Ambient Intelligence: Third International Joint Conference, AmI 2012, Pisa, Italy, November 13-15, 2012, Proceedings*, Lecture Notes in Computer Science, vol. 7683, Springer-Verlag Berlin Heidelberg, pp. 145–160. https://doi.org/10.1007/978-3-642-34898-3_10.

Singla J., 2015, *Comparative Study of Mamdani-Type and Sugeno-Type Fuzzy Inference Systems for Diagnosis of Diabetes*, [in:] *2015 International Conference on Advances in Computer Engineering and Applications*, pp. 517–522. https://doi.org/10.1109/ICACEA.2015.7164799.

Sousa M.J., Moutinho A., Almeida M., 2019, *Classification of potential fire outbreaks: A fuzzy modeling approach based on thermal images*, Expert Systems with Applications, vol. 129, pp. 216–232. https://doi.org/10.1016/j.eswa.2019.03.030.

Takagi T., Sugeno M., 1985, *Fuzzy identification of systems and its applications to modeling and control*, EEE Transactions on Systems, Man, and Cybernetics, vol. SMC-15, no. 1, pp. 116–132. https://doi.org/10.1109/TSMC.1985.6313399.

Wang L.X., Mendel J.M., 1992, *Generating Fuzzy Rules by Learning from Examples*, IEEE Transactions on Systems, Man, and Cybernetics, vol. 22, no. 6, pp. 1414–1427. https://doi.org/10.1109/21.199466.

Wolberg W.H., Street W.N., Mangasarian O.L., 1995, *Wisconsin Diagnostic Breast Cancer Database*, https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+%28 Diagnostic%29 [access: 12.09.2019].